WHAT IS CLAIMED IS:

1.   A document processor which displays and prints in a predetermined format a plurality of document data input thereto, comprising:

5   document memory which stores input document data;

         selection unit which selects all or part of document data stored in said documents memory;

         characteristics extraction unit which extracts data relating to characteristics of letter rows from all or part of

10   the document data selected by said selection unit;

         work processing unit which work-processes all or part of the document data based on the data relating to characteristics of letter rows extracted by said characteristics extraction unit; and

15   output unit which outputs all or part of the document data work-processed by said work processing unit.

2.   The document processor according to claim 1, wherein said output unit comprises item value set unit which sets a plurality

20   of item values based on the contents of all or part of the document data work-processed by said work-processing unit; and totalization unit which totalizes all or part of the document data for each item value set by said item value set unit;

         said output unit outputs all or part of the document data

25   in the format of a table having an item value as at least one

138

axis.

3.   The document processor according to claim 1, wherein said output unit outputs all or part of the document data work-processed by said work processing unit together with all or part of the document data in its state prior to work-processing by said work processing unit.

4.   The document processor according to claim 1, wherein said document memory further stores all or part of the document data work-processed by said work processing unit.

5.   The document processor according to claim 1, wherein said selection unit further selects all or part of the document data output by said output unit.

6.   The document processor according to claim 1, wherein said document memory further stores data relating to contents of the work processing.

7.   A document classification device which classifies documents based on contents thereof comprising:
       input unit which inputs document data;
       language analyzer unit which analyzes document data input by said input unit and obtains language analysis information;

139

vector creation unit which obtains document characteristic vectors for the document data based on the language analysis information obtained by said language analyzer unit;

5   classification unit which classifies documents based on the degree of similarity between document characteristic vectors created by said vector creation unit, and creating clusters of documents;

        cluster characteristics calculation unit which
10   calculates cluster characteristics, which are characteristics of clusters of documents created by said classification unit; and

        classification category memory which stores cluster characteristics, calculated by said cluster characteristics
15   calculation unit, as constituent elements of classification categories.

8.   A document classification device which classifies documents based on contents thereof comprising:

20   input unit which inputs a document data;

        language analyzer unit which analyzes document data input by said input unit and obtains language analysis information;

        vector creation unit which creates document characteristic vectors for the document data based on the
25   language analysis information obtained by said language

140

analyzer unit;

classification unit which classifies documents based on the degree of similarity between document characteristic vectors created by said vector creation unit, and creates clusters of documents;

cluster characteristics calculation unit which calculates cluster characteristics, which are characteristics of clusters of documents created by said classification unit;

display unit which displays the cluster characteristics calculated by said cluster characteristics calculation unit;

cluster selection specification unit which selects predetermined clusters from cluster of documents created by said classification unit; and

classification category memory which stores cluster characteristics, calculated by said cluster characteristics calculation unit, as constituent elements of classification categories.

9. The document classification device according to claim 8, further comprising document characteristic vector memory which stores document characteristic vectors created by vector creation unit; and

vector correction unit which corrects document characteristic vectors stored in said document characteristic vector memory, so that document characteristic vectors of

documents belonging to clusters selected by said cluster selection unit are deleted;

said classification unit which classifies documents based on the document characteristic vectors corrected by said vector correction unit.

10.   The document classification device according to claim 8, further comprising document characteristic vector memory which stores document characteristic vectors created by vector creation unit; and

document expression space correction unit which corrects document expression space when determining the degree of similarity between document characteristic vectors stored in said document characteristic vectors memory, based on a characteristics amount calculated from clusters selected by said cluster selection unit;

said classification unit classify the documents based on the degree of similarity between document characteristic vectors created by said vector creation unit, using the document expression space corrected by said document expression space correction unit.

11.   The document classification device according to claim 9, further comprising document characteristic vector memory which stores document characteristic vectors created by vector

142

creation unit; and

document expression space correction unit which corrects

the document expression space when determining the degree of

similarity between document characteristic vectors stored in

5   said document characteristic vectors memory, based on a

characteristics amount calculated from clusters selected by

said cluster selection unit;

said classification unit classify the documents based on

the degree of similarity between document characteristic

10 vectors created by said vector creation unit, using the document

expression space corrected by said document expression space

correction unit.

12.   The document classification device according to claim 8,

15 further comprising selection information appending unit which

appends selection information showing the fact of selection

when all or part of the documents belonging to a cluster of

documents created by said classification unit have been

selected;

20   said display unit displays the cluster characteristics,

and also displays the selection information appended by said

selection information appending unit.

13.   The document classification device according to claim 8,

25 wherein said classification category memory stores cluster

143

characteristics and/or information created by an operator, in addition to all or part of the documents belonging to a cluster of documents selected by said selection specification unit, as constituent elements of classification categories.

5

14. A document classification device which classifies document clusters in accordance with contents thereof comprising:

document input unit which inputs document data groups;

10   document dividing unit which divides document data into one or multiple divided document data based on a predetermined reference;

document-divided document map creation unit which creates a map showing the correspondence between the document

15   data and the divided document data;

divided document classification unit which classifies the divided document data;

divided document classification result creation unit which creates divided document classification result

20   information based on a classification result of said divided document classification unit; and

document classification result creation unit which creates classification result information of the above document data using the document-divided document map and the divided

25   document classification result information.

144

15. The document classification device according to claim 14, further comprising document save unit which saves the document data;

   divided document save unit which saves the divided document data; and

   document-divided document map save unit which saves a document-divided document map created by said document-divided document map creation unit.

16. The document classification device according to claim 15, further comprising divided document classification result save unit which saves the divided document classification result information created by said divided document classification result creation unit.

17. The document classification device according to claim 14, wherein a plurality of divided document data created by said document dividing unit comprises the document data in its state prior to being divided.

18. The document classification device according to claim 14, wherein said document dividing unit divides document data based on information relating to the structure of the document data.

19.   The document classification device according to claim 14, further comprising document element extraction unit which extracts elements in the document data;

element-accompanying information extraction unit which extracts element-accompanying information accompanying the elements extracted by said document element extraction unit;

said document dividing unit divides the document data using elements extracted by said document element extraction unit, or the elements and element-accompanying information extracted by said element-accompanying information extraction unit.

20.   The document classification device according to claim 14, wherein said document dividing unit divides the document data in compliance with a specified specification range.

21.   The document classification device according to claim 14, wherein said document dividing unit divides the document data based on the number of letters, the number of sentences, or both the number of letters and the number of sentences.

22.   The document classification device according to claim 14, wherein said document classification result creation unit extracts and presents information showing document data, and representative information accompanying the document data, as

146

classification result information.

23.   The document classification device according to claim 22, wherein said document classification result creation unit

5 extracts and presents information showing divided document data, and representative information accompanying the divided document data, as classification result information.

24.   A document processing method which outputs a plurality

10 of input document data in order to display or print the document data in a predetermined format, comprising the steps of:

storing input document data;

selecting all or part of the document data stored in the storing step;

15   extracting data relating to characteristics of letter rows from all or part of the document data selected in the selection step;

work-processing all or part of the document data based on the data relating to characteristics of letter rows extracted

20 in the characteristics extraction step; and

outputting all or part of the document data work-processed in the work-processing step.

25.   The document processing method according to claim 24,

25 wherein the step of outputting comprises the steps of setting

147

a plurality of item values based on the contents of all or part
of the document data work-processed in the work-processing
step; and totalizing all or part of the document data for each
item value set by in the item value setting step; and

5   outputs all or part of the document data in the format
of a table having an item value as at least one axis.

26.   The document processing method according to claim 24,
wherein the step of outputting comprises outputting all or part
10   of the document data work-processed in the work-processing step
together with all or part of the document data in its state prior
to work-processing in the work-processing step.

27.   The document processing method according to claim 24,
15   wherein the step of storing further comprises storing all or
part of the document data work-processed in the work-processing
step.

28.   The document processing method according to claim 24,
20   wherein the step of selecting further comprises selecting all
or part of the document data output in the output step.

29.   The document processing method according to claim 24,
wherein the step of storing a document further comprises storing
25   data relating to contents of the work processing.

148

30. A document classification method which classifies documents based on contents thereof comprising the steps of:

inputting a document data;

language-analyzing document data input in the step of inputting and obtaining language analysis information;

creating document characteristic vectors for the document data based on the language analysis information obtained in the step of language-analyzing;

classifying documents based on the degree of similarity between document characteristic vectors created in the step of creating vectors, and creating clusters of documents;

calculating cluster characteristics, being characteristics of clusters of documents created in the step of classifying; and

storing cluster characteristics, calculated in the step of calculating cluster characteristics, as constituent elements of classification categories.

31. A document classification method of classifying documents based on contents thereof, comprising the steps of:

inputting a document data;

language-analyzing document data input in the step of inputting and obtaining language analysis information;

creating document characteristic vectors for the

149

document data based on the language analysis information obtained in the step of language-analyzing;

classifying documents based on the degree of similarity between document characteristic vectors created in the step of

5   creating vectors, and creating clusters of documents;

calculating cluster characteristics, which are characteristics of clusters of documents created in the step of classifying;

displaying the cluster characteristics calculated in the

10   step of calculating cluster characteristics;

selecting predetermined clusters from cluster of documents created in the step of classifying; and

storing cluster characteristics, calculated in the step of calculating cluster characteristics, as constituent

15   elements of classification categories.

32.   The document classification method according to claim 31, further comprising the step of correcting document characteristic vectors stored in the step of storing document

20   characteristic vectors, so that document characteristic vectors of documents belonging to clusters selected by the step of selecting clusters are deleted;

the step of classifying comprising classifying documents based on the document characteristic vectors corrected by the

25   step of correcting vectors.

150

33.   The document classification method according to claim 31, further comprising a step of correcting document expression space when determining the degree of similarity between document characteristic vectors stored in the step of storing document characteristic vectors, based on a characteristics amount calculated from clusters selected in the step of selecting clusters;

the step of classifying comprising classifying documents based on the degree of similarity between document characteristic vectors created in the step of creating vectors, using the document expression space corrected in the step of correcting the document expression space.

34.   The document classification method according to claim 32, further comprising a step of correcting document expression space when determining the degree of similarity between document characteristic vectors stored in the step of storing document characteristic vectors, based on a characteristics amount calculated from clusters selected in the step of selecting clusters;

the step of classifying comprising classifying documents based on the degree of similarity between document characteristic vectors created in the step of creating vectors, using the document expression space corrected in the step of

151

correcting the document expression space.

35.   The document classification method according to claim 31, further comprising the steps of appending selection information

5   showing the fact of selection when all or part of the documents belonging to a cluster of documents created in the step of classifying have been selected;

the step of displaying comprising displaying the cluster characteristics, and also displaying the selection information

10   appended in the step of appending selection information.

36.   The document classification device according to claim 31, wherein the step of creating classification categories comprises creating cluster characteristics and/or information

15   created by an operator, in addition to all or part of the documents belonging to a cluster of documents selected in the step of specifying selection, as constituent elements of classification categories.

20   37.   A document classification method which classifies document clusters in accordance with contents thereof comprising the steps of:

inputting document data groups; dividing document data into one or multiple divided document data based on a

25   predetermined reference; creating a map showing the

152

correspondence between the document data and the divided document data; classifying the divided document data; creating divided document classification result information based on the classification result of classifying the divided documents; and

5 creating classification result information of the document data using the document-divided document map and the divided document classification result information.

38. A computer-readable recording medium in which is stored
10 programs for executing a document classification method, which document classification method comprising the steps of:

storing input document data;

selecting all or part of the document data stored in the storing step;

15 extracting data relating to characteristics of letter rows from all or part of the document data selected in the selection step;

work-processing all or part of the document data based on the data relating to characteristics of letter rows extracted
20 in the characteristics extraction step; and

outputting all or part of the document data work-processed in the work-processing step.

39. A computer-readable recording medium in which is stored
25 programs for executing a document classification method, which

153

document classification method comprising the steps of:

inputting a document data;

language-analyzing document data input in the step of inputting and obtaining language analysis information;

5   creating document characteristic vectors for the document data based on the language analysis information obtained in the step of language-analyzing;

classifying documents based on the degree of similarity between document characteristic vectors created in the step of

10   creating vectors, and creating clusters of documents;

calculating cluster characteristics, being characteristics of clusters of documents created in the step of classifying; and

storing cluster characteristics, calculated in the step

15   of calculating cluster characteristics, as constituent elements of classification categories.

40.   A computer-readable recording medium in which is stored programs for executing a document classification method, which

20   document classification method comprising the steps of:

inputting a document data;

language-analyzing document data input in the step of inputting and obtaining language analysis information;

creating document characteristic vectors for the

25   document data based on the language analysis information

obtained in the step of language-analyzing;

classifying documents based on the degree of similarity between document characteristic vectors created in the step of creating vectors, and creating clusters of documents;

5   calculating cluster characteristics, which are characteristics of clusters of documents created in the step of classifying;

displaying the cluster characteristics calculated in the step of calculating cluster characteristics;

10   selecting predetermined clusters from cluster of documents created in the step of classifying; and

storing cluster characteristics, calculated in the step of calculating cluster characteristics, as constituent elements of classification categories.

15

41.   A computer-readable recording medium in which is stored programs for executing a document classification method, which document classification method comprising the steps of:

inputting document data groups; dividing document data
20   into one or multiple divided document data based on a predetermined reference; creating a map showing the correspondence between the document data and the divided document data; classifying the divided document data; creating divided document classification result information based on the
25   classification result of classifying the divided documents; and

155

creating classification result information of the document data using the document-divided document map and the divided document classification result information.